Assessing the use of mode preference as a covariate for the estimation of measurement effects between modes: a sequential mixed mode experiment

Vandenplas, Caroline; Loosveldt, Geert; Vannieuwenhuyze, Jorre T.A.

methods, data, analyses | Vol. 10(2), 2016, pp. 119-142 DOI: 10.12758/mda.2016.011

Assessing the Use of Mode Preference as a Covariate for the Estimation of Measurement Effects between Modes. A Sequential Mixed Mode Experiment

Caroline Vandenplas[1], *Geert Loosveldt*[1] & *Jorre T. A. Vannieuwenhuyze*[2]

1 KU Leuven

2 Utrecht University

Abstract

Mixed mode surveys are presented as a solution to increasing survey costs and decreasing response rates. The disadvantage of such designs is the lack of control over mode effects and the interaction between selection and measurement effects. In a mixed mode survey, measurement effects can put into doubt data comparability between subgroups, or similarly between waves or rounds of a survey conducted using different modes. To understand the extent of measurement effects, selection and measurement effects between modes have to be disentangled. Almost all techniques to separate these effects depend on covariates that are assumed to be mode-insensitive and to fully explain selection effects. Most of the time, these covariates are sociodemographic variables that might be mode-insensitive, but fail to sufficiently explain selection effects. The aim of this research is to assess the performance of mode preference variables as covariates to evaluate selection and measurement effects between modes.

In 2012, a mixed mode survey – a web questionnaire followed by face-to-face interviews– was conducted alongside the face-to-face European Social Survey in Estonia (Ainsaar et al., 2013). The questionnaire included mode preference items. In this paper, the effects of the trade-offs between the two assumptions on the precision of estimated selection and measurement effects are compared. The results show that while adding the mode preference to the propensity score model seems to increase the explanatory power of web participation, it decreases the correlation between propensity scores and target variables. In addition, the estimated selection and measurement effects do not always fit the expectation that more selection effects are explained and more measurement effects are detected.

Keywords: mixed mode surveys; selection effects; measurement effects; mode preference; back-door method

1 Introduction

Mixed mode surveys are those during which different modes are offered simultaneously or sequentially. Such surveys have increased in popularity and are often implemented to adapt surveys to the needs or preferences of respondents. The implementation of mixed mode surveys is aimed at reducing costs, increasing response rates, and decreasing nonresponse bias, compared with traditional single mode surveys–especially face-to-face and telephone surveys. However, data collected using different modes may lead to differences in survey estimates due to mode effects. Mode effects can be separated into (1) selection effects, which are defined as differences in the responding sample due to different non-coverage or nonresponse errors between the modes, and (2) measurement effects, which occur when the answer from the same respondent would differ if a different data collection mode was used (Voogt & Saris, 2005; Weisberg, 2005).

1.1 Selection and Measurement Effects Between Modes

Selection effects between modes in a mixed mode survey can be desirable if they help to diversify the sample of respondents. Indeed, different modes may have different coverage problems and different levels of nonresponse bias (Dillman, Smyth, & Christian, 2009; de Leeuw, 2005). For example, the declining coverage of landline telephone surveys could be compensated for by adding a web questionnaire or face-to-face interviews for 'mobile only' individuals. Moreover, depending on their abilities and availability, individuals may be more likely to answer in one mode than in another. For example, web respondents are typically more likely to be higher educated and have a higher income, and are less likely to be elderly or from a minority compared with the general population. Indeed, people with these characteristics are more likely to be connected to the Internet, to use it frequently, and to have greater computer skills (Zillien & Hargittai, 2009; de Leeuw, 2005; Bimber, 2000). However, results concerning the benefits of using mixed mode surveys to reduce selection bias are mixed (e.g., Revilla, 2015; Medway & Fulton, 2012; Millar & Dillman, 2011; Holmberg et al., 2010; Smyth et al., 2010; US Census

Acknowledgements

We would like to thank the ESS Core Scientific Team for giving us access to the data, and the Estonian ESS team for the implementation of the European Social Survey mixed mode experiment in their country. We are also very grateful for the comments of two anonymous reviewers, which helped to improve the paper.

Direct correspondence to

Caroline Vandenplas, Center for Sociological Research, KU Leuven, Parkstraat, 45, 3000 Leuven, Belgium

E-mail: caroline.vandenplas@kuleuven.be

Bureau, 2010; Eva et al., 2010; Dillman, Phelps, et al., 2009; Gentry & Good, 2008; Fowler et al., 2002).

Measurement effects between modes may be problematic, especially if survey results need to be compared across rounds, across countries, or between subgroups in a country. When considering measurement effects, the measurement in one mode is often taken as the benchmark. Dillman (2000: chapter 6) points to differences in normative and cognitive consideration between modes, as well as interactions between the two. Especially when mixing interviewer-based and self-administered modes, the presence or absence of an interviewer and the aural or visual presentation of the items may lead to different stimuli and answering processes. The presence of an interviewer may increase socially desirable effects (the respondent taking social norms into consideration when answering the questions) and acquiescence (the tendency of the respondent to agree with the underlying statement of the question). Moreover, the visual presentation in a self-administered survey mode may increase primacy effects–choosing the first acceptable answer read–compared with aural presentation, which can favor recency effects–choosing the last acceptable answer heard. These effects can be reinforced by the lack of control over the cognitive efforts made by respondents in self-administered surveys, allowing them to not read the question and the answer options fully.

1.2 Back-door Method

Because the measured difference between alternative modes is a combination of selection and measurement effects, an important and complex issue is that of separating the two types of effects. To solve this confounding problem, Vannieuwenhuyze and colleagues (2010) suggest applying causal inference theory. In particular, the *back-door method* (Pearl, 2009; Morgan & Winship, 2009) can be applied to disentangle measurement and selection effects. The back-door method involves the inclusion of a set of variables X into the analysis model, where X explains the selection effects between different modes. The back-door method is based on two assumptions, the *mode selection ignorability assumption,* which requires that X fully captures the selection effects between the modes, and the *mode-insensitivity assumption*, which requires that the measurement of X is independent of the mode in which it is measured. Another proposed method to separate selection and measurement effects between modes is to re-interview respondents using another mode to estimate the measurement effects (Klausch, Hox, & Schouten, 2015; Klausch, Schouten, & Hox, 2015; Schouten et al., 2013).

Many existing attempts to separate selection effects from measurement effects in mixed mode surveys rely on the back-door method (e.g., Kolenikov & Kennedy, 2014; Vannieuwenhuyze et al., 2014; Vannieuwenhuyze & Loosveldt, 2013; Vannieuwenhuyze et al., 2012; Lugtig et al., 2011; Heerwegh & Loosveldt, 2011; Jäckle et al., 2010; Hayashi, 2007). However, most of these attempts are based on a set of sociodemographic variables that can be argued to be mode-insensitive, but probably fail to fully explain selection effects, i.e. to make the groups responding in different modes comparable. Therefore, variables that can complement sociodemographic variables as covariates in the back-door method should be found, of which one example may be mode preference variables.

1.3 Mode Preference

Mode preference reflects the fact that there may be different modes (Groves & Kahn, 1979) in which sampled people are more likely to answer (Olson et al., 2012; Shi & Fan, 2007; Miller et al., 2002). Based on this preference, mixed mode surveys are expected to have better response rates, because the choice of data collection mode theoretically increases the response propensity. For instance, some people feel uncomfortable with web questionnaires, because they are not familiar with using computers or the Internet, whereas others may perceive a web questionnaire as less intrusive than a face-to-face interview (Smyth et al., 2014). Mode preferences can therefore be hypothesized to be good predictors of the selected survey mode in a mixed mode survey (Olson et al., 2012) and can act as back-door variables. However, questions about mode preference may be subject to measurement effects. Previous research shows that respondents are more likely to endorse the mode they participate in, and therefore in which the mode preference is measured (Millar, O'Neil, & Dillman, 2009; Gesell, Drain, & Sullivan, 2007; Tarnai & Paxson, 2004; Groves & Kahn, 1979).

Although such variables are not expected to fulfill the mode-insensitivity assumption, they may offer a better trade-off between compliance with the mode-insensitivity assumption and compliance with the mode-selection ignorability assumption, compared with using sociodemographic variables when evaluating measurement and selection effects between modes in a mixed mode survey. Moreover, a possible solution to this mode-sensitivity is the creation of a latent variable that allows the control of measurement effects between modes, using a multi-group structural model. This requires, of course, at least three items measuring mode preference.

1.4 Different Sets of Covariates for the Back-door Method

To test the hypothesis that mode preference variables achieve a better balance between the two assumptions, we compare three sets of variables in this article: Only sociodemographic, sociodemographic combined with mode preferences, and sociodemographic combined with a latent mode preference variable. On the one hand, selection effects could be underestimated when only sociodemographic variables are included as back-door variables. As a consequence, the selection effects would not be completely corrected when applying the back-door method and the residual selection effects would be wrongly attributed to the measurement effect. The measurement effects estimates would then be biased: Over or under-estimated if the selection and measurement effects are in respectively the same or the opposite direction. On the other hand, selection effects might be estimated more accurately when variables about mode preferences are included, given the expected strong relationship between mode selection effects and the mode preference variables. However, the consequences of the mode-sensitive nature of mode preference variables on the estimated selection effects are difficult to predict. They could accentuate the selection effects and lead to an overcorrection of the selection effect when applying the back-door method. Conversely, the mode-sensitivity of the mode preference could result in an underestimation of the selection effects, or even introduce a completely random component. Lastly, the inclusion as a covariate of a latent mode preference variable built on three measurements of mode preferences should allow for a more-precise estimation of the selection effects. Indeed, the latent variable is independent of random measurement errors on the three specific measurements, and forcing the structural model to be the same in both modes should reduce measurement effects.

2 Data

The European Social Survey (ESS) is an academically-driven survey, designed to study the interactions between changing institutions, attitudes, beliefs, and behavioral patterns in Europe. The ESS started in 2002 and has been repeated every two years. Since its first round, great efforts have been made within the ESS to collect high quality data, and to ensure cross-national and cross-cultural comparability. Given the issues of the increasing costs of face-to-face surveys and declining response rates in some countries, it was decided to explore the possibility of mixed mode survey designs as an alternative to the traditional face-to-face interviews.

In 2012, a mixed mode survey was conducted in Estonia in parallel to round six of the main ESS survey. A simple random sample of 925 individuals was drawn from the population register to participate in a sequential, mixed mode survey, involving a web questionnaire (mode *a*), followed by a face-to-face phase (mode *b*) for the sample units who did not participate in the web component. A first invitation letter to the web survey containing a hyperlink and an individual password was sent to the 925 sampled individuals on 18 September. Two reminders (copies of the invitation letter) were sent respectively two weeks and four weeks after the first invitation letter was sent, as well as a last reminder to people who started the online questionnaire without completing it within approximatively five weeks. On

22 October, the face-to-face stage started for all the sample units who had not completed the web questionnaire. In the end, 356 people (38.4%) responded via the web survey and 230 (24.8%) completed the face-to-face interview, making a total of 586 respondents. The final response rate of 63.3% is not significantly different from the response rate for the main ESS survey (2380 out of 3702 = 64.2%, Chi square p = 0.7), where response rates are calculated as the number of completed interviews/questionnaires divided by the sample size, ignoring ineligible people.

An analysis of characteristics reveals some differences between the web and face-to-face respondents in the mixed mode survey. Results show that web respondents on average were younger, higher educated, and more likely to live in the North of Estonia compared with the face-to-face respondents (Ainsaar et al., 2013).

In addition to the usual ESS questionnaire, the mixed mode survey included questions about mode preference, survey attitudes, and the perceived accuracy of the survey.

The questionnaire contains three mode preference related variables that are considered as possible auxiliary variables to control for selection effects between the web and the face-to-face component of the survey. These variables are:

- Web participation (RPWEB): In general, how often would you respond to surveys like this one if you were invited to complete an internet questionnaire?
- Phone participation (RPPHONE): In general, how often would you respond to surveys like this one if you were invited to complete a telephone interview?
- Face-to-face participation (RPF2F): In general, how often would you respond to surveys like this one if you were invited to complete a face-to-face interview?

The answer categories are: 1= never, 2 = once in a while, 3 = about half of the time, 4 = most of the time, 5 = always. In the hope of reducing measurement effects between the modes, the variables related to mode preferences did not directly ask about the preferred mode, but were instead designed so that the mode preference could be deduced from them.

Item nonresponse to mode preference variables reduced the responding sample from 582 to 556. As a consequence, all analyses are performed considering these 556 respondents.

Table 1 displays the means and standard deviations of the three mode-preference variables among web respondents and among face-to-face respondents. The mean for ‘phone participation’ is similar between the two groups but the means for ‘web participation’ or ‘face-to-face participation’ are very different. As expected, web respondents have a higher mean for ‘web participation’ than face-to-face respondents, and face-to-face respondents have a higher mean for ‘face-to-face participation’ than web respondents.

Given the categorical nature of the variables, we also show the distribution of these variables in Figure 1.

Table 1 Means and standard deviations of the variables about mode preferences for web respondents and for face-to-face respondents

Variables	Web mean	Web standard deviation	Face-to-face mean	Face-to-face standard deviation
Web	2.50	1.19	1.60	0.97
Phone	1.56	0.95	1.71	1.06
Face-to-face	1.71	1.06	2.90	1.28

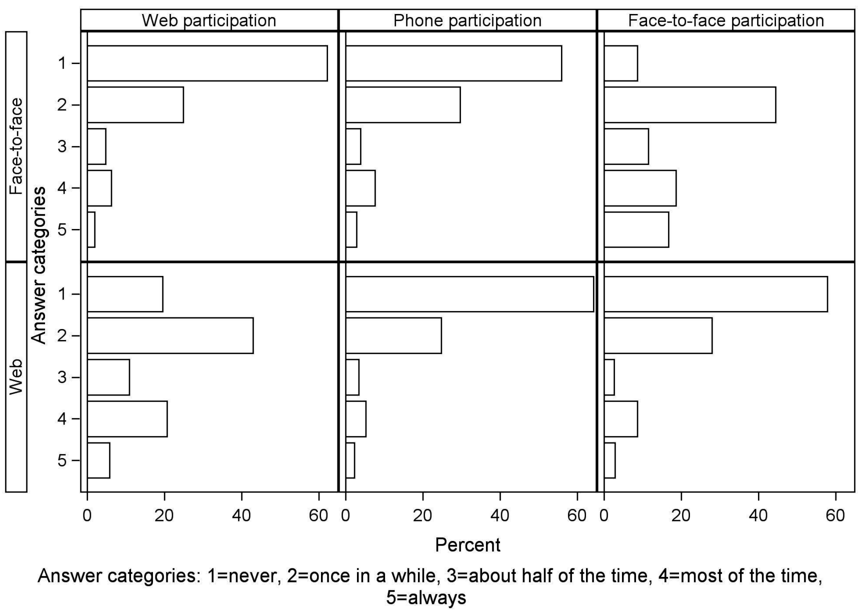

Figure 1 Distibution of the mode preference variables

We also need a set of substantive survey variables (Y) that could suffer from mode effects. We first consider measurement and selection effects on four items about survey attitudes. Although these variables were not part of the standard ESS questionnaire, but were added in the mixed mode version of the ESS in round 6, we examine these items as we expect them to suffer from strong measurement and selection effects between the web and the face-to-face mode. Indeed, these items are known to be subject to social desirability effects (negative measurement effects) (Vannieuwenhuyze et al., 2013). Moreover, the web respondents are also believed to have a more positive attitude toward surveys (positive selection effects) because they were 'early' respondents who did not require the face-to-face follow-up to

Table 2 Means and standard deviations of the variables about attitudes toward surveys for web respondents and for face-to-face respondents

Variable	Web mean	Web standard deviation	Face-to face mean	Face-to-face standard deviation
Privacy	5.11	3.10	6.51	2.96
Trust	5.19	2.43	6.35	2.64
Interest	4.39	3.08	6.28	2.75
Usefulness	6.20	2.60	6.91	2.48

participate. Therefore, we consider these 'attitude toward surveys' variables as test variables.

- Privacy (PRVCY): Do you find that surveys are an invasion of people's privacy? with the answer categories from 0 = A complete invasion of private life, to 10 = No invasion of private life at all (inverted compared with the original).
- Trust (TRSTSVY): Do you trust results obtained from a survey like this? with answer categories from 0 = No trust at all, to 10 = Complete trust.
- Interest (INTSVY): Do you find surveys like this interesting? with answer categories from 0 = Not interesting at all, to 10 = Completely interesting.
- Usefulness (USFLSVY): Do you find surveys like this useful? with answer categories from 0 = Not useful at all, to 10 = Completely useful.

Table 2 shows the means and standard deviations of these variables for the web and the face-to-face respondent groups. As expected from the social desirability hypothesis, the face-to-face respondent's means are higher than those for the web respondents.

We then consider three, four-point scale items related to attitudes toward immigration. The hypothesis for these variables is that web respondents have more positive attitudes (positive selection effects). Indeed, web respondents are in general higher educated, which is usually associated with a more positive attitude toward immigration. Moreover, the web respondents are expected to give more positive answers (positive measurement effects) due to a primacy effect caused by the vertical display of the answers in the web questionnaire, the answer category 'allow some' being read before 'allow few'. These variables are:

- Same ethnicity (IMSMETN): To what extent do you think Estonia should allow people of the same race or ethnic group as most Estonian people to come and live here?
- Different ethnicity (IMDFETN): How about people of a different race or ethnic group from most Estonian people?

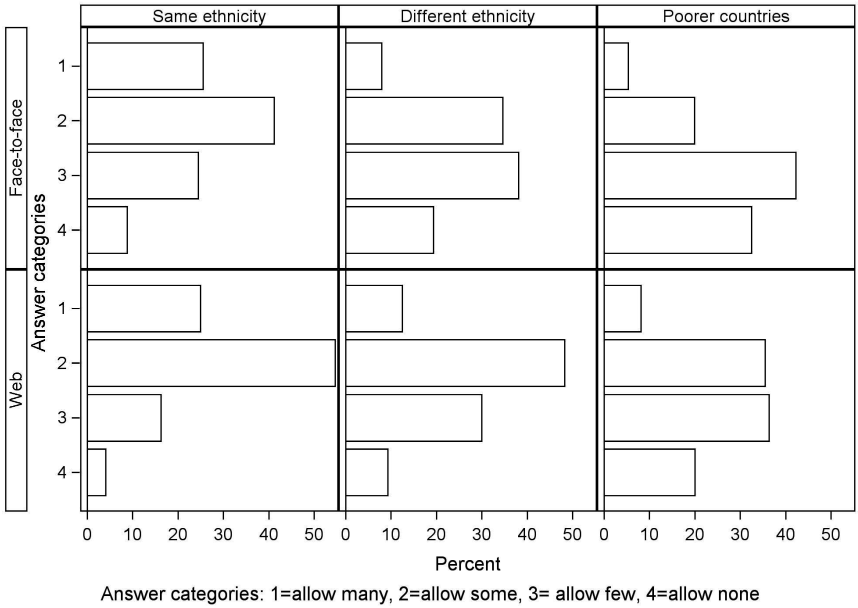

Figure 2 Distribution of the frequency of the chosen category for web respondents and for face-to-face respondents

- Poorer country (IMPRCNTR): How about people from the poorer countries outside Europe?

The answer categories are: 1= allow none, 2 = allow a few, 3 = allow some, 4 = allow many (inverted compared with the original scale).

Figure 2 shows the frequency of each answer category for these variables among web respondents and among face-to-face respondents. In this figure, the original negative scale is displayed where 1 = allow many, 2 = allow some, 3 = allow a few, and 4 = allow none. From the figure, it is clear that the category '2 = allow some' is more frequently chosen than the category '3 = allow a few' in the web questionnaire compared with the face-to-face interview.

Lastly, another set of three variables about attitudes toward immigration that have an 11-point scale rather than a four-point scale are considered.

- Economy (IMBGECO): Would you say it is generally bad or good for Estonia's economy that people come to live here from other countries? with answer categories from 0 = Bad for the economy, to 10 = Good for the economy.
- Culture (IMUECLT): And, using this card, would you say that Estonia's cultural life is generally undermined or enriched by people coming to live here from other countries? with answer categories from 0 = Cultural life is undermined, to 10 = Cultural life is enriched.

Table 3 Means and standard deviations of the variables about attitudes toward immigration for web respondents and for face-to-face respondents

Variable	Web mean	Web standard deviation	Face-to-face mean	Face-to-face standard deviation
Immigration and economy	4.98	2.30	4.66	2.52
Immigration and culture	5.45	2.53	5.43	2.44
Immigration and country	4.70	2.13	4.48	2.32

- Country (IMWBCNT): Is Estonia made a worse or a better place to live by people coming to live here from other countries? with answer categories from 0 = A worse place to live, to 10 = A better place to live.

The hypotheses for these variables are again that the web respondents have more positive attitudes (positive selection effects) toward immigration, but that the answers of the face-to-face respondents are more inclined to suffer from social desirability (negative measurement effect).

Table 3 displays the means and standard deviations of these variables for the web and for the face-to-face respondent groups. The web respondents' means are higher than those of the face-to-face respondents, despite the expected effect of social desirability, but supporting the hypothesized positive selection effect for the web.

3 Methods

The aim behind disentangling the two types of mode effects–selection and measurement–is to correct measurement effects so that results are comparable across rounds or waves of a repeated survey or across subgroups in one round. In the studied mixed mode design, the web is considered as the principal mode. Consequently, the answers given in the face-to-face interviews (observed answers) should be corrected so that they become equivalent to the answers that would have been given in a web questionnaire (counterfactual answers). To do this, we apply the back-door method, wherein a set of auxiliary variables (X) is used to model the selection effect. In the first step, the web (mode *a)* responding group is matched with the face-to-face (mode *b*) responding group through, for example, weighting. This means that the web respondents are given a weight such that the weighted web respondent group is equivalent to the face-to-face responding group, typically if considering the distribution of the set of auxiliary variables X. In the second step, the difference in estimates between the web and face-to-face respondent is split into

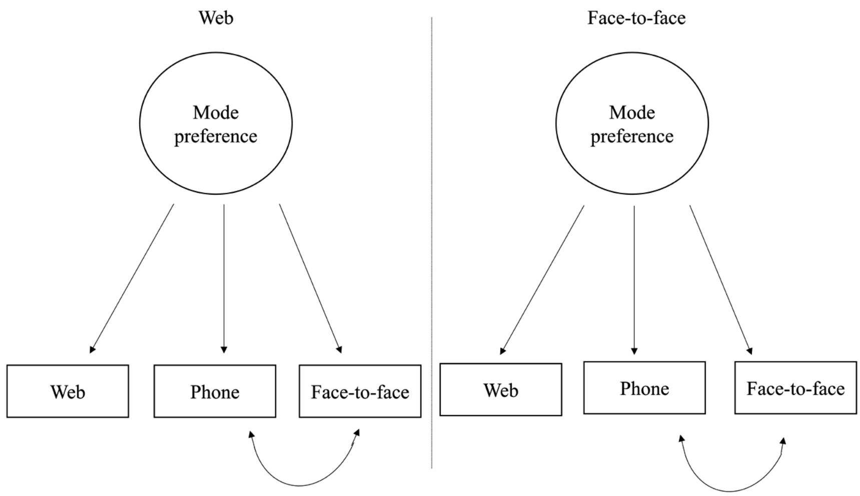

Figure 3 The mode preference latent model

(1) selection effects estimated by the difference between the web estimates and the weighted web estimates and (2) the measurement effects estimated as the difference between the weighted web estimates and the face-to-face estimates. The accuracy of the estimated measurement and selection effects depends on the compliance of the covariates (X) to the mode-insensitivity and the mode-selection ignorability assumptions.

3.1 Latent Mode Preference Variable

Given that the compliance of the set of covariates (X) with the mode-insensitivity assumption can be doubted when the mode preference variables are introduced, we create a mode preference latent variable based on the three mode-preference related variables. The construction of the latent variable should allow us to control for the measurement effect on specific items, while still extracting the essence of mode preference. A multi-group structural equation approach is applied, where the groups are defined by the modes. This approach allows us to construct equivalent latent measurement models in both modes. Because full scale equivalence between the modes appears to be too strong a requirement (CFI = 0.106, RMSEA = 0.490, and SRMR = 0.235), the equality of intercept for 'face-to-face participation' (RPF2F) and 'web participation' (RPWEB) is relaxed, but the metric equivalence and the intercept equality for the 'phone participation' (RPPHONE) are retained. The 'face-to-face participation' and 'web participation' are more likely to be subject to measurement effects between the modes than 'phone participation'. A correlation between 'phone participation' and 'face-to-face participation' is also allowed, in order to improve the model fit (CFI = 0.993, RMSEA = 0.089, and SRMR = 0.028). This seems theoretically acceptable, as both types of data collection modes involve an interaction with an interviewer. We used the lavaan package in R to create this measurement model.

3.2 Propensity Score Weighting

We apply propensity score weighting to correct for the selection effect between the web group and the face-to-face group. The propensity score of respondent *i* is defined as the probability of *i* to participate in the web mode (mode *a*), given a set of (back-door) variables $x_1(i), x_2(i), \ldots, x_j(i)$ estimated by the logistic model (Lee & Valliant, 2008): $logit(p(\boldsymbol{x})) = \beta_0 + \beta_1 x_1 + \ldots + \beta_j x_j$.

Once estimated, in line with Lee (2006), the propensity scores are ordered and partitioned into *K* strata of equal size. We use ten strata (deciles) following the strategy shown in recent literature (Matsuo et al., 2010; Loosveldt & Sonck, 2008; Schonlau et al., 2009). If n_k denotes the total number of respondents in stratum *k*, $n_{k,b}$ the number of respondents in stratum *k* responding by a face-to-face interview (mode *b*), and $n_{k,a}$ the number of respondents in stratum *k* responding by web (mode *a*), the adjustment factor for all web respondents (mode *a*) in stratum *k* is then defined as $w_k = \frac{n_{k,b} / n_b}{n_{k,a} / n_a}$ where n_a is the total number of web respondents and n_b the total number of face-to-face respondents. This weighting scheme equates the (weighted) proportion of web respondents in stratum *k* with the proportion of face-to-face respondent in stratum *k*. The weighted number of web respondents in stratum *k* is given by $n_{k,a} w_k = n_{k,a} \frac{n_{k,b} / n_b}{n_{k,a} / n_a} = n_{k,b} \frac{n_a}{n_b} = n_a \frac{n_{k,b}}{n_b}$ and hence, the weighted proportion of web respondents is $\frac{n_{k,b}}{n_b}$.

3.3 The Propensity Models Based on the three Sets of Variables Considered

We calculate three sets of propensity scores in order to assess the efficiency of three different sets of back-door variables *X*.

In the first step, we used sociodemographic variables to calculate propensity scores and their associated rank strata. These variables are gender, age (4 categories: 15-29, 30-44, 45-64, and 65+), education (lower-secondary or less, upper-secondary, post-secondary or tertiary, and bachelor's, master's, or doctorate), work (in paid work or not), and geographical region of residence. In the logistic regression to estimate the propensity to participate in the web component, only education and

Table 4 Number of web and face-to-face respondents in each deciles of the propensity score distribution depending on the considered set of auxiliary variables

	Web			Face-to-face		
Deciles	Socio-demo.	+ three mode pref.	+ latent mode pref.	Socio-demo.	+ three mode pref.	+ latent mode pref.
0	11	1	10	45	54	45
1	28	3	30	27	53	26
2	30	17	27	27	39	29
3	34	31	34	22	24	21
4	38	39	35	18	17	21
5	34	45	37	22	11	19
6	38	46	39	16	9	16
7	44	53	45	11	2	12
8	40	57	44	14	0	11
9	50	55	46	7	0	9

age significantly contributed to the model. Nonetheless, all variables were retained in the logistic model in line with Lee and Valliant (2008: p. 178). The analyses were repeated with only significantly contributing variables, without implications for the results or the conclusions.

In the second step, the three mode preference variables were included, from which only 'web participation' and 'face-to-face participation' significantly contributed to the model. When mode preference variables were included, the two last strata – with the highest propensities to participate to the web component – did not contain any of the face-to-face respondents. For this reason, the web respondents in these strata (n=55+57) were given a weight of 0. This is a violation of the overlap assumption of propensity score matching methodology, which states that every unit should have a non-zero probability to be attributed to any of the groups (modes). This represents a limitation of our analysis.

In the third step, in an attempt to control for possible measurement effects on the three mode-preference variables, these three variables were replaced by a mode preference latent variable in the logistic model.

Table 4 displays the number of web respondents and of face-to-face respondents in each rank stratum for the three sets of auxiliary variables. The rank strata are deciles that were created after the web and face-to-face respondents had been ordered by propensity scores.

The distributions of web and face-to-face respondents over the propensity deciles are quite similar when these deciles are based on the sets including only sociodemographic variables and sociodemographic variables together with the mode preference latent variable. By contrast, the distribution over the deciles of web and face-to-face respondents are different when these deciles are based on the set including sociodemographic variables and the three mode preference related variables. In this case, there are almost no web respondents in the first deciles and no face-to-face respondents in the last deciles.

3.4 Estimating Selection and Measurement Effects

Assuming that the variables X are mode-insensitive and entirely explain the selection effect, the selection effects and the measurement effects can be expressed as follows. The answer given by respondent i in mode m, which is either web, a, or face-to-face, b, to a particular item (survey attitude or attitude toward immigration) is denoted $y_{i,m}$.

Taking the sum of the web (mode a) respondents, the selection effect is calculated as the difference before and after weighting:

$$
\begin{aligned}
S_a\left(\mu(Y)\right) &= \frac{\sum_{i=1}^{n_a} y_{i,a}}{n_a} - \frac{\sum_{i=1}^{n_a} w_{k,i} y_{i,a}}{n_a} = \frac{\sum_{i=1}^{n_a} y_{i,a}}{n_a} - \frac{\sum_{i=1}^{n_a} \frac{n_{k,b}/n_b}{n_{k,a}/n_a} y_{i,a}}{n_a} \\
&= \frac{\sum_{i=1}^{n_a} y_{i,a}}{n_a} - \frac{n_a \sum_{i=1}^{n_a} \frac{n_{k,b}}{n_{k,a}} y_{i,a}}{n_a n_b} = \frac{\sum_{k=1}^{10} n_{k,a}\mu_{k,a}}{n_a} - \frac{\sum_{k=1}^{10} n_{k,b} \sum_{i=1}^{n_{k,a}} \frac{y_{i,a}}{n_{k,a}}}{n_b} \\
&= \frac{\sum_{k=1}^{10} n_{k,a}\mu_{k,a}}{n_a} - \frac{\sum_{k=1}^{10} n_{k,b}\mu_{k,a}}{n_b},
\end{aligned}
$$

where $\mu_{k,a}$ is defined as the mean of web respondents over stratum k.

It should be noted that these selection effects only concern whether respondents participate in the web component of the survey rather than in the face-to-face interviews. Non-respondents are not considered.

Taking the sum of the web (mode a) respondents and face-to-face (mode b) respondents, the measurement effect is calculated as the difference of weighted responses in the web respondent group measured in mode a (web) and the (unweighted) responses in the face-to-face respondent group measured in mode b (face-to-face):

$$
M_a\left(\mu(Y)\right) = \frac{\sum_{i=1}^{n_a} w_{k,i} y_{i,a}}{n_a} - \frac{\sum_{i=1}^{n_b} y_{i,b}}{n_b} = \frac{\sum_{k=1}^{10} n_{k,b}\mu_{k,a}}{n_b} - \frac{\sum_{k=1}^{10} n_{k,b}\mu_{k,b}}{n_b},
$$

where $\mu_{k,a} / \mu_{k,b}$ is defined as the mean of web/face-to-face respondents over stratum k.

3.5 Significance of the Selection and Measurement Effects

Because the propensity scores are based on the respondent sample and not the full population, there is a certain degree of sampling error associated with their estimation. To integrate this level of variability, we used the bootstrap method (Efron, 1979) with 500 replicates. This means that we resampled the responding sample with a replacement 500 times – so that the replicated sample is the same size as the original responding sample–and performed the full analysis, from calculating the propensity scores to estimating the measurement and selection effects for each replicate. The variance and standard error of this collection of 500 estimates of selection and measurement effects (assuming a normal distribution) are estimates of the variability of the estimated effects. The significance of the selection and measurement effects are based on these estimated standard errors.

4 Results

Our aim in this paper was to assess the performance of mode preference variables to control for selection effects, with the goal of estimating measurement effects in the face-to-face component compared with the web component in a sequential mixed mode survey.

4.1 Model Fit of the Propensity Models

Because mode preference variables are expected to better explain selection effects between the modes, propensity models including mode preference as the independent variable should be more appropriate to predict the selected mode, and should therefore lead to a better fit of the propensity model. This better fit is confirmed by the ESS data when including the three mode preference variables alongside the sociodemographic variables: The model fit strongly improves (AIC goes from 724.5 to 420.6, pseudo-R from 0.18 to 0.66). This improvement is significant according to the residual Chi-square test (Score: 140.95/69.72 with p-values <0.001 for 4 degrees of freedom for face-to-face participation and web participation respectively). These results confirm our expectation concerning the relevance of these mode preference variables. Nevertheless, including the mode-preference latent variables instead of the three raw variables does not lead to an improvement of the model fit.

The difference in model fit improvement when using the three mode-preference variables or when using the corresponding latent variable might be an indication that the strong relationship between mode preference and the mode of participation may be explained by a violation of the mode-insensitivity assumption. Because of measurement effects on the mode-preference variables themselves, the relationship between mode preference and mode of participation may be highly overestimated.

4.2 Correlation of Propensity Scores with Target Variables

Ideal weighting variables should not only correlate with the propensity to participate in the web component of the survey, but also with the target variables (Groves, 2006; Little & Vartivarian, 2003, 2005; Kalton & Flores-Cervantes, 2003; Kalton & Maligalig, 1991; Little, 1986). As our estimated propensity scores were used to construct our weighting strata, Spearman correlation coefficients were estimated between the different target variables (attitudes toward surveys and toward immigration) and the propensity scores based on different sets of covariates (Table 5).

When considering the propensity scores based on the three mode-preference variables, results yield reduced correlation between the propensity score and the target variables. Hence, even though the mode-preference variables improve the propensity model fit of the logistic propensity model, they reduce the strength of the correlation with target variables.

When considering the propensity scores based on the latent mode-preference variable, results yield similar correlations between the propensity score and the target variables compared with when considering the propensity score based only on sociodemographic variables: Slightly stronger for attitudes toward surveys and lower for attitudes toward immigration when the mode-preference latent variable is added.

Looking at the sign of the correlations, unexpected negative correlations between the propensity score and the attitudes toward surveys (privacy, trust, interest, and usefulness) should be noted. Indeed, as web respondents are in general higher educated, and furthermore, ‘early’ respondents in the sequential mixed mode surveys, we expect them to have more positive attitudes toward surveys. Hence, we expect a higher propensity to participate in the web survey to be positively correlated with survey attitudes, and not negatively. A possible explanation for this surprising result is measurement effects on the surveys attitude variables causing face-to-face respondents to give more positive answers due to the presence of an interviewer.

4.3 Estimation of Measurement and Selection Effects

The effect of including mode-preference items in the propensity model to detect selection and measurement effects is central to our paper. Table 6 shows the

Table 5 Spearman correlations between target variables and propensity score for the different propensity models

Variables	sociodemo.	+ three mode pref.	+ latent mode pref.
Privacy	-0.12 ***	-0.09 *	-0.18 ***
Trust	-0.09 *	-0.09 *	-0.12 ***
Interest	-0.16 ***	-0.11 **	-0.23 ***
Usefulness	-0.12 ***	-0.05	-0.19 ***
Same ethnicity	0.07 *	0.05	0.05
Different ethnicity	0.24 ***	0.22 ***	0.22 ***
Poorer countries	0.27 ***	0.15 ***	0.24 ***
Economy	0.18 ***	0.10 *	0.17 ***
Culture	0.08 *	0.07 ‘	0.07 *
Country	0.15 ***	0.12 **	0.15 ***

‘p<0.1, * p<0.05, ** p<0.01 and ***p<0.001.

unweighted means for the web respondents, the weighted means for web respondents when the three different sets of auxiliary variables are included in the propensity model, and the mean for the face-to-face respondents. The last six columns in Table 6 display the selection and measurement effects estimates using the three different sets of covariates.

The results in Table 6 partially confirm our hypothesis concerning the direction of the selection effects, which was expected to be positive for all the variables of interest. The selection effects are indeed positive, or in most cases, not significantly different from 0 (α=0.05), independent of the set of auxiliary variables. The only exception is ‘same ethnicity’ when the three mode-preference variables are included in the propensity model, which displays a negative selection effect.

Moreover, the hypothesis concerning the measurement effect on the variables of interest is also supported by the results in Table 6. The measurement effects are all negative, or not significantly different from 0, for the 11-point scale variables about attitudes toward surveys and toward immigration. Moreover the measurement effects are positive or not significantly different from 0 for the 4-point scale variables about attitudes toward immigration.

Lastly, adding the three mode-preference variables does not lead to larger positive selection effect estimates than when only considering the sociodemographic variables. By contrast, some of the positive selection effects detected with sociodemographic variables only become not significantly different from 0. Furthermore, the selection effect on ‘same ethnicity’ is estimated as negative when the three mode-preference variables are added. When the latent variable ‘mode preference’ is added to the propensity model, the estimated selection effects are similar

Table 6 Selection and measurement effects between web and face-to-face respondents when using different auxiliary variables in the propensity score model

Variable name	Web mean	Weighted web mean			Face-to-face mean	Sociodemo.		+ three mode pref.		+ latent mode pref.	
		Sociodemo.	+ three mode pref.	+ latent mode pref.		Sel.	Meas.	Sel.	Meas.	Sel.	Meas.
Privacy	5.11	5.02	6.31	5.18	6.51	0.09	-1.50*	-1.19	-0.21	-0.07	-1.33**
Trust	5.19	5.24	5.16	5.28	6.35	-0.04	-1.11*	0.03	-1.18*	-0.09	-1.06**
Interest	4.39	4.44	4.10	4.67	6.28	-0.05	-1.84*	0.29	-2.18*	-0.28	-1.61**
Usefulness	6.21	6.23	7.14	6.43	6.91	-0.02	-0.68*	-0.93	0.23	-0.22	-0.48
Same ethnicity	3.01	2.93	3.43	2.98	2.83	0.07	0.10	-0.42*	0.59*	0.03	0.15
Different ethnicity	2.64	2.46	2.43	2.49	2.31	0.18*	0.15	0.21	0.11	0.15*	0.17
Poor countries	2.32	2.07	2.35	2.13	1.98	0.24*	0.09	-0.03	0.37	0.18*	0.15
Economy	4.98	4.55	4.86	4.59	4.67	0.43*	-0.12	0.12	0.19	0.39*	-0.08
Culture	5.45	5.08	4.88	5.33	5.43	0.38	-0.36	0.57	-0.55	0.13	-0.11
Country	4.70	4.28	4.12	4.41	4.48	0.42*	-0.20	0.58	-0.36	0.29	-0.07

*p<0.05

to selection effects estimated when only sociodemographic variables are considered in the propensity mode. These results are not in line with our expectations that the inclusion of mode preference in the propensity model would help to detect larger positive selection effects. There is no real pattern in the influence on selection effects of introducing the three variables concerning mode preference, showing that the measurement effects on these variables interfere greatly with the estimation of the propensity scores. The 'true' selection effects of the web compared with the face-to-face component are, however, unknown. Moreover, the overlap assumption of the propensity methodology is violated here, some web respondents could not be matched to face-to-face respondents, which could have unexpected consequences. Therefore, we are limited in drawing conclusions about the performance of the mode preference as a covariate to estimate selection effects.

5 Conclusion

The aim of this research was to test whether the inclusion of mode-preference variables in a set of covariates to control for selection effects between survey modes would offer a better trade-off between compliance with the mode-selection ignorability assumption and compliance with the mode-insensitivity assumption. To draw conclusions on the usability of mode-preference variables, three set of covariates–(1) only sociodemographic variables, (2) adding three mode-related variables, and (3) adding a mode-preference latent variable–were used in a propensity score model to evaluate the participation of respondents in the web component of a mixed-mode survey. The resulting selection and measurement effects were then compared.

The main finding is that there is no evidence that including mode-preference variables in the sets of covariates leads to more accurate estimates of the selection effects. Two cases can be distinguished: (1) no pattern can be found in the consequences for the estimated selection effects of adding the three mode-preference related variables, not controlling for mode effects on these variables, and (2) estimated selection effects are not larger (in the presumed direction) when adding the latent mode-preference variable that was constructed to control for measurement effects on the mode-preference measurements. The violation of the mode-sensitivity assumptions by the mode-preference variables seems to cause an irreversible problem, leading to the non-usability of these variables as covariates in the back-door method. Moreover, the attempt to cancel the mode-sensitivity of the mode-preference variables by the construction of a latent variable wiped out the impact of the mode-preference variables on the selection effects.

We should mention some limitations of this research. First, empirical evidence of the absence of the added value of mode preference as a covariate is limited by the relatively small sample size and by the particularities of the survey examined: Restricted to Estonia, comparing only two modes offered sequentially, and not appointed randomly. A second limitation is that the 'true' selection effects are unknown. A third limitation is the violation of the overlap assumption of propensity matching methodology when the three mode variables are added, which could affect our conclusions. More research, in an experimental context, may be necessary to generalize our findings.

Furthermore, this research highlights the presence of measurement effects between modes in different aspects. Although almost no significant measurement effect was found on the means of the variables reflecting attitudes toward immigration, large measurement effects were found on the variables reflecting attitudes toward surveys. Therefore, attitudes toward surveys are clearly the most sensitive to social desirability. Even if adding the mode-preference variables separately reduced some of the estimated measurement effects, taking the latent variables into consideration increased them again. To conclude, even if the measurement effects between the modes are probably overestimated, the present study supports their presence.

These findings point to the risk of comparing results between data collection modes. A lot remains unexplained about the answering processes of respondents in different modes and their effects on measurement error. A possible solution would be the unimode design, in which items are designed to be robust across modes (Dillman, 2000: chapter 6).

Finally, more research might be needed in order to find adequate covariates to control for selection and measurement effects between survey modes, and to study differences in response styles between modes depending on question designs.

References

Ainsaar, M., Lilleoja, L., Lumiste, K., & Roots, A. (2013). *ESS Mixed Mode Experiment Results in Estonia (CAWI and CAPI Mode Sequential Design).* Tartu: University of Tartu, ISBN 978-9985-4-0757-8.

Bimber, B. (2003). Measuring the Gender Gap on the Internet. *Social Science Quarterly*, 81(3), 1-11.

de Leeuw, E. D. (2005). To Mix or not to Mix Data Collection Mode in Surveys. *Journal of Official Statistics,* 21(2), 233-255.

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method.* New York: John Wiley.

Dillman, D. A., Smyth, J.D., & Christian, L.M. (2009). *Internet, mail and mixed-mode surveys: the tailored design method* (3rd Ed). Hoboken: Wiley.

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, J. (2009). Response Rate and Measurement Differences in Mixed-Mode Surveys using Mail, Telephone, Interactive Voice Response (IVR) and the Internet. *Social Science Research,* 38, 1-18.

Efron, B. (1979). Bootstrap Method: Another Look at the Jackknife. *The Annals of Statistics,* 7(1), 1-26.

Eva, G., Loosveldt, G., Lynn, P., Martin, P., Revilla M., Saris W., & Vannieuwenhuyze, J. (2010). *Assessing the Cost-Effectiveness of Different Modes for ESS Data Collection*. London: City University.

Fowler, F. J., Gallagher, P. M. , Stringfellow, V. L., Zaslavsky, A. M, Thompson, J. W., & Cleary, P. D. (2002). Using Telephone Interviews to Reduce Nonresponse Bias to Mail Surveys of Health Plan Members. *Medical Care,* 40, 190-200.

Gentry, R., & Good, C. (2008). *Offering Respondents a Choice of Survey Mode: Use Patterns of an Internet Response Option in a Mail Survey*. Paper presented at the Annual Conference of the American Association for Public Opinion Research, New Orleans. May 15-18, 2008.

Gesell, S. B., Drain, M., & Sullivan, M. P. (2007). Test of a web and Paper Employee Satisfaction Survey: Comparison of Respondents and Non-Respondents. *International Journal of Internet Science,* 2, 45-58.

Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly,* 70(5), 646-675.

Groves, R.M., & Kahn, R. (1979). *Survey by Telephone: a National Comparison with Personal Interviews.* New York: John Wiley and Sons.

Hayashi, T. (2007). The Possibility of Mixed-mode Surveys in Sociological Studies. *International Journal of Japanese Sociology,* 16, 51-63.

Heerwegh, D., & Loosveldt, G. (2011). Assessing Mode Effects in a National Crime Victimization Survey Using Structural Equation Models: Social Desirability Bias and Acquiescence. *Journal of Official Statistics,* 27, 49-63.

Holmberg, A., Lorenc, B., & Werner, P. (2010). Contact Strategies to Improve Participation via the web in a Mixed-Mode Mail and Web Survey. *Journal of Official Statistics,* 26(3), 465-480.

Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the Effect of Data Collection Mode on Measurement. *International Statistics Review,* 78(1), 3-20.

Kalton, G., & Flores-Cevantes, I. (2003). Weighting Methods. *Journal of Official Statistics,* 19 (1), 81-97.

Kalton, G., & Maligalig, D. (1991). A Comparison of Methods of Weighting Adjustment for Nonresponse. In *Proceedings of the 1991 Annual Research Conference* (pp. 401-428). Washington DC: U.S. Bureau of the Census.

Klausch, T., Hox, J., & Schouten, B. (2015). Selection error in single- and mixed mode surveys of the Dutch general population. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 178(4), 945-961. http://doi.org/10.1111/rssa.12102

Klausch, T., Schouten, B., & Hox, J. J. (2015). Evaluating Bias of Sequential Mixed-mode Designs against Benchmark Surveys. *Sociological Methods & Research*. http://doi.org/10.1177/0049124115585362.

Kolenikov, S., & Kennedy, C. (2014). Evaluating Three Approaches to Statistically Adjust for Mode Effects. *Journal of Survey Statistics and Methodology,* 2, 126-158.

Lee, S. (2006). Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel web Surveys. *Journal of Official Statistics,* 22(2), 329-349.

Lee, S., & Valliant, R. (2008). Weighting Telephone Samples using Propensity Scores. In J. Lepkowski et al. (Eds.), *Advances in Telephone Survey Methodology* (pp. 170-183). Hoboken: John Wiley and Sons.

Little, R. J. A. (1986). Survey Nonresponse Adjustments for Estimates of Means. *International Statistical Review*, 54(2), 139-157.

Little, R. J., & Vartivarian, S. (2003). On Weighting the Rates in Non-response Weights. *Statistics in Medicine,* 22(9), 1589-1599.

Little, R. J., & Vartivarian, S. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means? *Survey Methodology,* 31, 161-168.

Loosveldt, G., & Sonck, N. (2008). An Evaluation of the Weighting Procedure for an Online Access Panel Survey. *Survey Research Method,* 2(2), 93-105.

Lugtig, P. J., Lensvelt-Mulders, G. J. L. M., Frerichs, R., & Greven, F. (2011). Estimating Nonresponse Bias and Mode Effects in a Mixed-Mode Survey. *International Journal of Market Research*, 53(5), 669-686.

Matsuo, H., Billiet, J., Loosveldt, G., Berglund, F., & Kleven, Ø. (2010). Measurement and Adjustment of Non-Response Bias based on Non-Response Surveys : the Case of Belgium and Norway in the European Social Survey Round 3. *Survey Research Methods*, 4(3), 165-178. Retrieved from http://w4.ub.uni-konstanz.de/srm/article/viewFile/3774/4332

Medway, R. L., & Fulton, J. (2012). When More Gets You Less: A Meta-Analysis of the Effect of Concurrent Web Options on Mail Survey Response Rates. *Public Opinion Quarterly*, 76(4), 733-746.

Millar, M. M., O'Neill A. C., & Dillman, D. A. (2009). *Are mode Preferences Real? Technical Report 09-003.* Pullman WA: Social & Economic Sciences Research Center. Washington State University.

Millar, M. M., & Dillman, D. A. (2011). Improving Response to web and Mixed-Mode Surveys. *Public Opinion Quarterly,* 75(2), 249-269.

Miller, T. I., Miller-Kobayashi, M., Caldwell, E., Thurston. S., & Collett, B. (2002). Citizen Surveys on the Web: General Population Survey of Community Opinion. *Social Science Computer Review,* 20, 124-136.

Morgan, S. L., & Winship, C. (2009). *Counterfactuals and causal inference: methods and principles for social research.* New York: Cambridge University Press.

Olson, K., Smyth, J. D., & Wood, H. M. (2012). Does giving People their Preferred Survey Mode actually Increase Survey Participation? *Public Opinion Quarterly,*76(4), 611-635.

Pearl, J. (2009). *Causality: Model Reasoning and Interference (2nd Ed.*). New York: Cambridge University Press.

Revilla, M. (2015). Comparison of the Quality Estimates in a Mixed-Mode and a Unimode Design: an Experiment from the European Social Survey. *Quality and Quantity,* 49, 121-1238.

Schonlau, M., van Soest, A., Kapteyn, A., & Couper, M. (2009). Selection Bias in Web Surveys and the Use of Propensity Scores. *Sociological Methods & Research*, 37(3), 291-318. http://doi.org/10.1177/0049124108327128

Schouten, B., van den Brakel, J., Buelens, B., van der Laan, J., & Klausch, T. (2013). Disentangling mode-specific selection and measurement bias in social surveys. *Social Science Research,* 42(6), 1555-1570. http://doi.org/10.1016/j.ssresearch.2013.07.005

Shi, T.-H., & Fan, X. (2007). Response Rate and Mode Preferences in Web-Mail Mixed-Mode Surveys: a Meta-Analysis. *International Journal of Internet Science*, 2(1), 59-82.

Smyth, J. D., Olson, K., & Millar, M. M. (2014). Identifying Predictors of Survey Mode Preference. *Social Science Research,* 48, 135-144.

Smyth, J. D., Dillman D. A., Christian, L. M., & O'Neill A. C. (2010). Using the Internet to Survey Small Towns and Communities: Limitations and Possibilities in the Early 21st Century. *American Behavioral Scientist,* 53, 1423-1448.

Tarnai, J., & Paxson, M. C. (2004). Survey Mode Preference of Business Respondents. In *Proceedings of the Survey Research Section of the American Statistical Association* (pp. 4866-4872).

US Census Bureau. (2010). Design and Methodology: American Community Survey, Washington DC: US Census Bureau. Available at: https://www.census.gov/content/dam/Census/library/publications/2010/acs/acs_design_methodology.pdf

Vannieuwenhuyze, J., Loosveldt, G., & Molenberghs, G. (2010). A Method for Evaluating Mode Effects in Mixed-Mode Surveys. *Public Opinion Quarterly,* 74(5), 1027-1045.

Vannieuwenhuyze, J., Loosveldt G., & Molenberghs, G. (2012). A Method to Evaluate Mode Effects on the Mean and Variance of a Continuous Variable in Mixed-Mode Surveys. *International Statistical Review,* 80, 306-322.

Vannieuwenhuyze, J., & Loosveldt, G. (2013). Evaluating Relative Mode Effects in Mixed-Mode Surveys: Three Methods to Disentangle Selection and Measurement Effect. *Sociological Methods and Research,* 42(1), 82-104.

Vannieuwenhuyze, J., Loosveldt, G., & Molenberghs, G. (2014). Evaluating Mode Effects in Mixed-Mode Survey Data Using Covariate Adjustment Models. *Journal of Official Statistics*, 30(1), 1-21.

Voogt, R. J. J., & Saris, W. (2005). Mixed-Mode Designs: Finding the Balance between Nonresponse Bias and Mode Effects. *Journal of Official Statistics,* 21(3), 367-387.

Weisberg, H. F. (2005). *The Total Survey Error Approach: a Guide to the New Science of Survey Research.* Chicago (III): University of Chicago.

Zillien, N., & Hargittai, E. (2009). Digital Distinction: Status-Specific Types of Internet Usage. *Social Science Quarterly*, 90(2), 274-291.